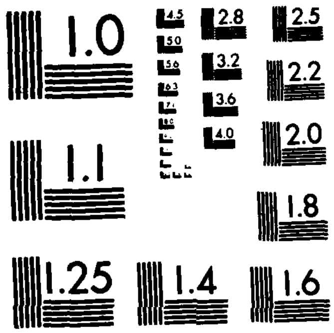
AD-A137 960 SOME CONSIDERATIONS IN ESTIMATING DATA TRANSFORMATIONS (U) WISCONSIN UNIV-MADISON MATHEMATICS RESEARCH CENTER G E BOX ET AL. DEC 83 MRC-TSR-2609 DAAG29-80-C-0041 1/1

UNCLASSIFIED F/G 12/1 NL

END
FILMED
3-84
DTIC



MICROCOPY RESOLUTION TEST CHART
NATIONAL BUREAU OF STANDARDS-1963-A

3

AD A137960

MRC Technical Summary Report #2609

SOME CONSIDERATIONS IN ESTIMATING DATA TRANSFORMATIONS

George E. P. Box and Conrad A. Fung

Mathematics Research Center
University of Wisconsin—Madison
610 Walnut Street
Madison, Wisconsin 53705

December 1983

(Received July 20, 1982)

Approved for public release
Distribution unlimited

Sponsored by

U. S. Army Research Office
P. O. Box 12211
Research Triangle Park
North Carolina 27709

UNIVERSITY OF WISCONSIN-MADISON
MATHEMATICS RESEARCH CENTER

SOME CONSIDERATIONS IN ESTIMATING DATA TRANSFORMATIONS

George E. P. Box and Conrad A. Fung*

Technical Summary Report #2609
December 1983

ABSTRACT

In a recent paper Bickel and Doksum claimed that procedures proposed by Box and Cox for estimating a transformation can be costly and unstable. ~~We~~ the document considers how the supposed cost and instability arise and illustrate ~~our~~ the authors' points by further analysis of textile data from the original paper. The analysis is used to make the further point that the cost of <u>not</u> making a transformation, when such is appropriate, can be extremely high. Some common sense advice on transformation analysis is given.

AMS (MOS) Subject Classifications: 62-07, 62A10, 62A15

Key Words: Transformations, Box-Cox, Bickel-Doksum, Efficiency, Stability

Work Unit Number 4 (Statistics and Probability)

---

*Current address: E. I. DuPont de Nemours and Co., Inc., Wilmington, DE 19898.

Sponsored by the United States Army under Contract No. DAAG29-80-C-0041.

SIGNIFICANCE AND EXPLANATION

Much statistical analysis is concerned with empirical models such as those commonly used in the analysis of variance and in regression analysis. Such models often suppose that the dependence of the response y upon the experimental conditions may be represented by some kind of simple model, with errors (independently) distributed with constant variance in normal distributions. It is sometimes true that such models can be rendered much more representationally adequate by making some transformation of the response y such as $\log y$, $y^{1/2}$ or y^{-1}. All the transformations just listed may be regarded as special cases of a power transformation class conveniently written as $y^{(\lambda)} = (y^{\lambda} - 1)/\lambda$. More generally we can define $y^{(\underset{\sim}{\lambda})}$ as referring to any class of transformations depending on parameters $\underset{\sim}{\lambda}$. Box and Cox (1964) proposed a method for estimating the parameters $\underset{\sim}{\lambda}$ for such a class of transformations at the same time that the model was fitted. The method has been widely used with considerable success.

In the above we use the term <u>empirical model</u> to mean one whose form does not rest on physical or mechanistic justification. It is not uncommon to tacitly imbue empirical models with more authority than they can sustain. For example, when variables are quantitative, empirical models such as polynomials, are most aptly regarded as useful mathematical "french curves" which can be adapted to graduate a variety of smooth functions by suitable adjustment of parameters among which, with co-equal status, are the transformation parameters $\underset{\sim}{\lambda}$.

Recently a theoretical paper by Bickel and Doksum (1981) concluded that these Box-Cox procedures could be "costly" and "highly unstable". We believe these conclusions are misleading. The present paper attempts to set out some of the issues and to supply some common sense advice on the use of these techniques.

The responsibility for the wording and views expressed in this descriptive summary lies with MRC, and not with the authors of this report.

SOME CONSIDERATIONS IN ESTIMATING DATA TRANSFORMATIONS

George E. P. Box and Conrad A. Fung*

SUMMARY

In a recent paper Bickel and Doksum (1981) concluded that a procedure suggested by Box and Cox (1964) for estimating a transformation can be costly and highly unstable. In a brief reply by Box and Cox (1982) argued that these conclusions result (a) from an inadmissible comparison of parameters in transformed models, (b) from neglect of the Jacobian of the transformation, and (c) from application of the procedure to problems where almost no information exists about the transformation parameter.

The present authors believe that, when appropriately used, the estimation of transformations is an extremely potent tool. We fear that the statement of conclusions by Bickel and Doksum may result in unfortunate misunderstandings. The purpose of this paper, therefore, is to attempt to clarify the issues more fully and further to make the point that to fail to transform appropriately can be costly.

1. SOME ISSUES

An elementary example

To concentrate ideas, imagine an empirical study of the relationship between y, the time in hours to death of an animal treated with a lethal

*Current address: E. I. DuPont de Nemours and Co., Inc., Wilmington, DE 19898.

Sponsored by the United States Army under Contract No. DAAG29-80-C-0041.

drug, and x, the dose of the drug in cc's. Suppose it is hoped to relate y to x by a simple empirical model of the form

$$y^{(\lambda)} = \alpha^{(\lambda)} + \theta^{(\lambda)}x + \varepsilon \tag{1}$$

where $y^{(\lambda)}$ is a power transformation* conveniently written as

$$y^{(\lambda)} = \begin{cases} (y^{\lambda} - 1)/\lambda & (\lambda \neq 0) \\ \log y & (\lambda = 0) \end{cases} \tag{2}$$

and where for some suitable choice of λ the errors ε are independently and approximately normally distributed with fixed variance σ^2. The bracketed superscripts λ on α and on θ are here introduced to indicate that these coefficients are measured in a scale which depends on that of $y^{(\lambda)}$. In practice it seems that useful power transformations mostly occur within the range $(-1 \leq \lambda \leq 1)$.

Units of $\theta^{(\lambda)}$

Suppose $\lambda = 1$. Then $\theta^{(1)}$ measures the rate at which time to death increases with dosage and the units of measurement are hours per cc of drug. By contrast, for $\lambda = -1$, $\theta^{(-1)}$ measures how the rate of dying decreases with dosage and the units of measurement are hours^{-1} per cc of drug. Obviously $\theta^{(1)}$ and $\theta^{(-1)}$ are incommensurable and hence not directly comparable. More generally it is essential to remember that when such transformations are applied, the units of measurement of the parameters and their meaning (as well as their numerical values) can alter dramatically as the transformation is changed.

*We use the power transformation for illustration. However, the procedures suggested by Box and Cox are in principle applicable to any class of non-linear parameter transformations including multidimensional $\underset{\sim}{\lambda}$ and for these, similar arguments would be likely to apply.

The near-linearity of power transformations over short ranges

As is well known, when the proportional change in y over the range of the data, as measured, for example, by y_{max}/y_{min}, is sufficiently small, power transformations will be nearly linear. For example, for data covering the range $y = 995$ to $y = 1005$, even for a power transformation as extreme as the reciprocal, a linear approximation of the form

$$y^{(\lambda)} = \omega_0 + \omega_1 y \tag{3}$$

where ω_0 and ω_1 are suitably chosen coefficients has a maximum error of less than 0.003%.

Now, much statistical analysis is invariant under linear recoding. In fact such recoding is often recommended routinely to increase clarity and for computational convenience; for such examples therefore, over a wide range of values of λ, the particular power transformation chosen (since it is essentially linear), may be a matter of indifference. Equivalently it will not matter whether any transformation is made at all. Now, the ability to estimate λ from data arises from the nonlinearity which the transformation induces. Thus, there may be very little information in the data about the parameter λ when y_{max}/y_{min} is small. Consequently, as is pointed out in elementary texts (for example, in Box, et al., (1978)), attempts to estimate λ in such situations may be fruitless. Of course information about λ depends on other factors besides y_{max}/y_{min}; in particular, on the coefficient of varia-tion of the data, on the design, and on the size of the sample. We intend to discuss these matters in a later paper. For the present we refer to data which allows only very imprecise estimation of λ as non-informative for λ.

2. CRITICISMS

Bickel and Doksum ignore the fact that $\theta^{(\lambda)}$'s are incommensurate for different values of λ. Let us temporarily do the same. Unless the data happen to be located and scaled so that they cluster about unity, because of the change in the units of measurement of $\theta^{(\lambda)}$ as λ is changed, the magnitude of $\theta^{(\lambda)}$ is typically highly dependent upon λ. Furthermore, this dependence can be increased without limit simply by changing the units in which y is measured, for example from hours to seconds. Thus, inevitably the distribution of the estimators $\hat{\lambda}$ and $\hat{\theta}^{(\lambda)}$ will also be highly dependent.

The "Cost" of estimating λ

The dependence induced by incommensurate scaling can therefore produce a marginal variance for $\hat{\theta}^{(\lambda)}$ that is greatly inflated in comparison with the conditional (λ-known) variance. It is this effect that accounts for most of what Bickel and Doksum call the cost of estimating λ.

The "instability" of Box-Cox procedures

For data which are non-informative for λ the marginal variance of $\hat{\lambda}$ will be very large ensuring that different samples of data generated by the same model can give very different estimates of λ which because of the high dependence between $\hat{\lambda}$ and $\hat{\theta}^{(\lambda)}$ will in turn produce very different estimates of $\theta^{(\lambda)}$ (in incommensurate units).

It is this effect which Bickel and Doksum refer to as the instability of Box-Cox procedures*.

*Of course as has been pointed out by Carroll and Ruppert (1980) "instability" is not transmitted into the estimates of the response. This parallels the effect found in linear regression where high correlation between parameter estimates and hence instability of the estimates induced by near-collinearity of the regressors is not transmitted into the estimates of the response for the region where the data are available.

The effect of not ignoring the Jacobian of the transformation

To the extent that the transformation $y^{(\lambda)}$ can be approximated by a linear function of y, the Jacobian of the transformation from y to $y^{(\lambda)}$ explains the change in scale. While Bickel and Doksum considered only the $y^{(\lambda)}$ form of the transformation of equation (2) Box and Cox took account of this linear scale change by employing, when the scale was important, the alternative form

$$z^{(\lambda)} = \begin{cases} (y^{\lambda} - 1)/(\lambda \dot{y}^{(\lambda-1)}) & (\lambda \neq 0) \\ \dot{y} \log y & (\lambda = 0) \end{cases} \tag{4}$$

where $\dot{y}$ is the geometric mean of the data. It will be observed that $z^{(\lambda)}$ is scaled in units of y whatever the value of λ. The factor $\dot{y}^{(\lambda-1)}$ from the Jacobian thus standardizes the scale of the transformed data $z^{(\lambda)}$ so that the coefficients $\theta^{(\lambda)}$ for analyses conducted in terms of $z^{(\lambda)}$, are more nearly comparable as λ varies. Conversely this factor measures the linear dependence between λ and θ induced by the arbitrary choice of the scale of y. Thus, suppose, for an example in which y was measured in tons, it happened that there was little change in the magnitude of $\hat{\theta}_{\lambda}$ as λ varied over a range from -1 to 1. Then for y measured in pounds, over the same range of λ, $\hat{\theta}_{\lambda}$ would vary over a range of $2{,}000^2 = 4{,}000{,}000$. It is true that this $z^{(\lambda)}$ form cures only the gross linear dependence between $\hat{\lambda}$ and $\hat{\theta}^{(\lambda)}$ and that residual non-linear dependence remains, but the effects complained of mostly occur when y_{max}/y_{min} is small so that $y^{(\lambda)}$ is almost linear in y. (See also Hinkley and Runger 1983.)

3. FURTHER EXAMINATION OF THE TEXTILE DATA

To illustrate these points we re-examine an example used by Box and Cox (1964). They describe how the data had been obtained by two textile scientists who had used a 3^3 factorial design in the study of a testing machine. The three input variables were x_1 = length of specimen, x_2 = amplitude, and x_3 = load. To this data the scientists had fitted a full 10-coefficient second degree equation

$$y = \theta + \Sigma\, \theta_i x_i + \sum_{i>j}\Sigma\, \theta_{ij} x_i x_j + \varepsilon \tag{5}$$

which gave complicated and messy conclusions. Box and Cox argued that general physical considerations suggested a log transformation in terms of which the model might simplify, and showed that for this data (for which y_{max}/y_{min} was over 40), λ could be accurately estimated and indeed lay very close to 0. With the log transformation $Y = \log y$, an excellent fit was obtained from a response equation of only first degree. Thus with the log metric all second order terms θ_{ij} in (5) could be omitted, and the response equation contained only 4 coefficients $(\theta, \theta_1, \theta_2, \theta_3)$. We here extend somewhat the Box-Cox Bayesian analysis* noting that the conclusions readily translate to a sampling theory context.

For illustration Figure 1 shows contours for the joint marginal posterior distributions of λ and one of the regression parameters θ_1. In Figure 1(a) the analysis is conducted in terms of $y^{(\lambda)}$ and in Figure 1(b) in terms of $z^{(\lambda)}$. Note in particular the great reduction in dependence resulting from the use of $z^{(\lambda)}$ rather than the $y^{(\lambda)}$ form.

Figures 2(a) and 2(b) enable us to study the nature of the model simplification arising from the log transformation for analyses conducted in terms of $y^{(\lambda)}$ and $z^{(\lambda)}$, respectively. They show the dependence on λ of

---

*In which, a priori, $p(\boldsymbol{\theta}, \log \sigma | \lambda) \propto \dot{y}^{k(\lambda-1)}$ and $p(\lambda)$ is locally uniform.

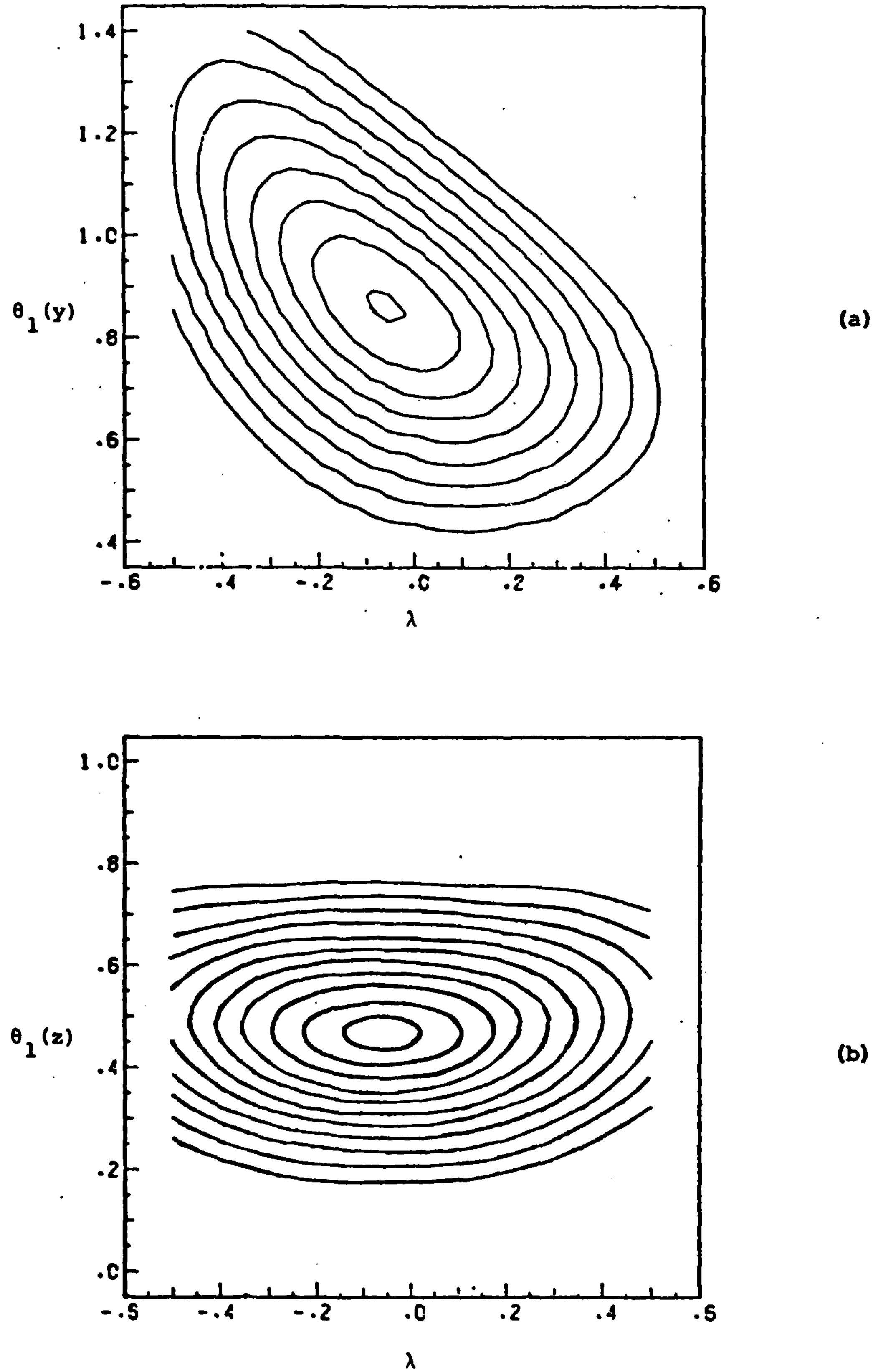


Figure 1. Contours of the joint marginal posterior for λ and θ_1,
(a) analysis in terms of $y^{(\lambda)}$
(b) analysis in terms of $z^{(\lambda)}$

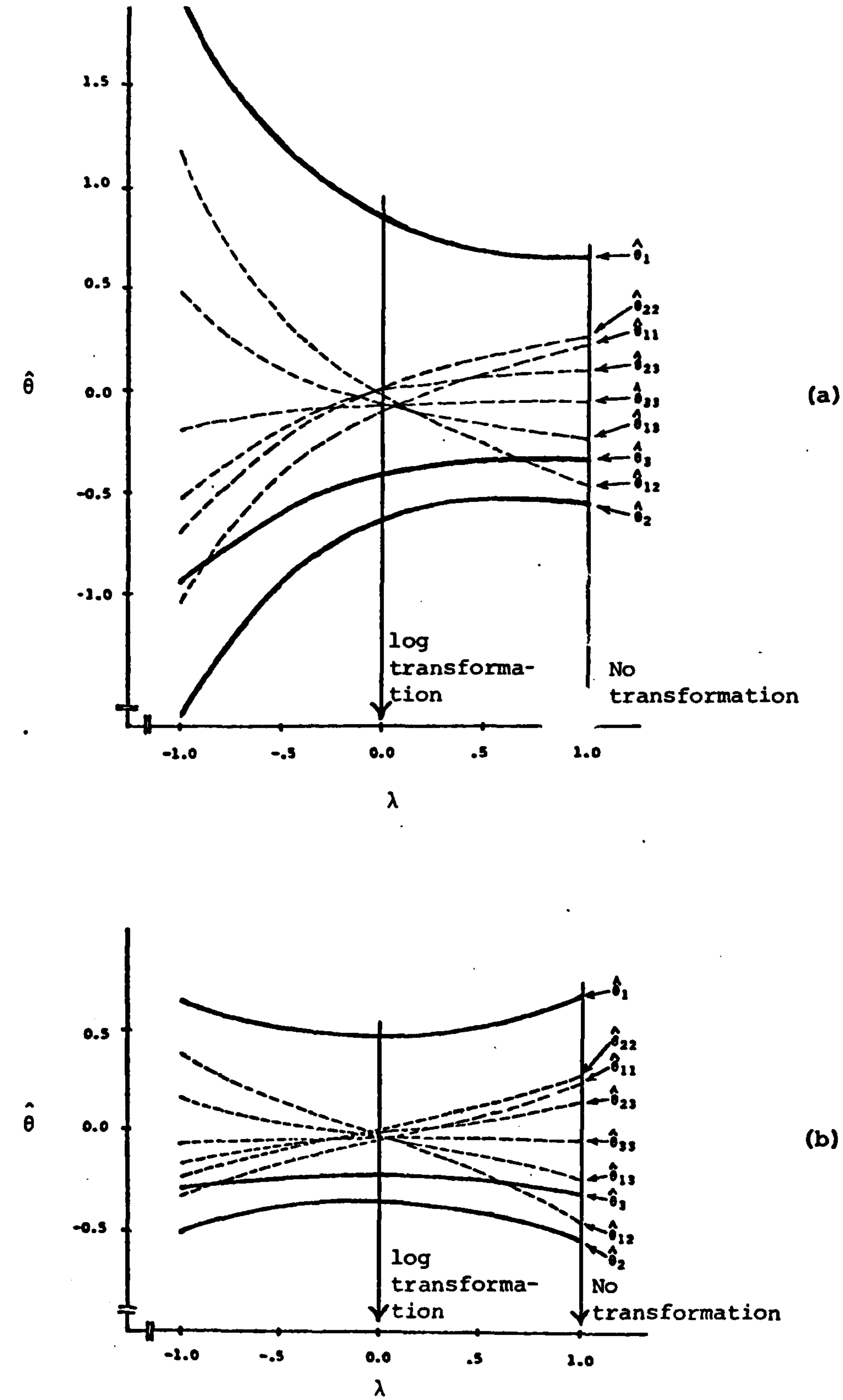


Figure 2. The dependence of estimated coefficients $\hat{\theta}$ on λ, with second order model
(a) analysis in terms of $y^{(\lambda)}$
(b) analysis in terms of $z^{(\lambda)}$

the conditional estimates $\hat{\underset{\sim}{\theta}}^{(\lambda)}$ (conditional Bayesian means) for the coefficients in a full second degree equation model. From either diagram it will be seen that with no transformation, $\lambda = 1$, (and more generally with values of λ not close to zero) nearly all the coefficients $\underset{\sim}{\theta}$ tend to be distinct from zero, making necessary a second degree approximation. By contrast, as we approach the log transformation, $\lambda = 0$, all second order terms become small, making relevant the much simpler first degree 4-parameter model.

Comparison of Figures 2(a) and 2(b) illustrates how the linear part of the dependence between λ and the estimated <u>first order</u> coefficients $\hat{\theta}_i^{(\lambda)}$ is eliminated by employing $z^{(\lambda)}$ instead of $y^{(\lambda)}$ (i.e. by taking account of the Jacobian). By contrast, notice how the strong dependence between λ and the <u>second order</u> coefficients $\hat{\theta}_{ij}^{(\lambda)}$ arising from inadequacy of the linear model when λ departs from zero is still evident in the $z^{(\lambda)}$ plot. In fact, as was otherwise demonstrated by Box and Cox, this dependence is clearly a major contributor to the remarkably precise estimation of λ possible in this particular example.

It is important to realize that, except for the purpose of illustrating our point, it makes little sense to study plots like Figures 1(a) and 2(a) for which the units of the vertical scale are meaningless. The dependence shown in these two diagrams is indeed rather moderate because in the units (10 hours) arbitrarily adopted in tabling this data the mean happens to be not very far from unity. Analysis in terms of hours, minutes or seconds would produce much more dramatic dependence.

4. THE COST OF NOT TRANSFORMING THE TEXTILE DATA

The original analysis of the textile data conducted by the textile scientists with no transformation (as if it were known that $\lambda = 1$) led to a model which was unnecessarily complicated. In addition it is possible to show that it also led to a great loss in efficiency. Comparison of the F-ratios calculated by Box and Cox for the original quadratic analysis and the linear analysis in the log scale, shows that the F-ratio for the latter is 7.6 times as large as that for the former. It is not entirely clear, however, what conclusion we should draw from this comparison, since the assumptions which justify the F distribution would be seriously invalid for untransformed data.

A more relevant assessment can be made as follows. From the careful earlier analysis of these data, it seemed that to a reasonable approximation, the linear model in $Y = \log y$ satisfied the standard assumptions, and we will assume this to be so. We further suppose that the object is to estimate y, and we will measure how well this is done by calculating the variances of the estimated responses taken over the 27 points at which the data were collected. Thus we are regarding the design points as sampling the space over the region where we might legitimately use the fitted equation.

In what follows, the standard normal theory assumptions which would justify the first-order model in the logged data are referred to as assumptions A_1, while the corresponding assumptions which would justify a second order model in the unlogged data are referred to as assumptions A_2.

In an obvious matrix notation, we can write the first order model in $Y = \ell n y$ fitted by least squares on the assumptions A_1, as

$$\hat{\underset{\sim}{Y}}_{(1)} = \underset{\sim}{X}_1 \hat{\underset{\sim}{\theta}}_{(1)} = \underset{\sim}{R}_1 \underset{\sim}{Y} \quad \text{with} \quad s_Y^2 = 0.0345$$

where

$$\underset{\sim}{R}_1 = \underset{\sim}{X}_1(\underset{\sim}{X}_1'\underset{\sim}{X}_1)^{-1}\underset{\sim}{X}_1' \; .$$

Correspondingly the quadratic model in y fitted on assumptions A_2 can be written

$$\hat{\underset{\sim}{y}}_{(2)} = \underset{\sim}{X}_2\hat{\underset{\sim}{\theta}}_{(2)} = \underset{\sim}{R}_2\underset{\sim}{y} \quad \text{with} \quad s_y^2 = .07392 \; ,$$

where

$$\underset{\sim}{R}_2 = \underset{\sim}{X}_2(\underset{\sim}{X}_2'\underset{\sim}{X}_2)^{-1}\underset{\sim}{X}_2' \; .$$

In the above s_Y^2 and s_y^2 represent the residual mean squares based respectively on $27 - 4 = 23$ and $27 - 10 = 17$ degrees of freedom.

Table (1) shows the results of the following calculations:

The "perceived" variances for the $\hat{y}_{(2)}$'s shown in the first column of the table are estimated on the (false) assumptions A_2 that the second order model in the untransformed response y is appropriate. They are the diagonal elements of the matrix $\underset{\sim}{R}_2 s_y^2$.

The "actual" variances for the $\hat{y}_{(2)}$'s given in the second column of the table were obtained using the following approximation. Let $\underset{\sim}{\Sigma}_y$ be the $n \times n$ covariance matrix of the y's. Then the variances of the $\hat{y}_{(2)}$'s are the diagonal elements of the matrix $\underset{\sim}{R}_2\underset{\sim}{\Sigma}_y\underset{\sim}{R}_2$. On the assumptions A_1, which we believe to be approximately correct, Σ_y is a 27×27 diagonal matrix whose i^{th} diagonal element is approximated by $\{\hat{y}_{(1)i}\}^2 s_Y^2$ where $\hat{y}_{(1)i} = \exp(\hat{Y}_{(1)i})$ is the estimated response obtained by taking the antilog of $\hat{Y}_{(1)i}$ which is fitted with the model we believe to be true.

The "attainable" variances shown in the third column are calculated as follows:

The variance $\mathrm{var}(\hat{Y}_{(1)i})$ of an estimated response $\hat{Y}_{(1)i}$ is the i^{th} diagonal element of the matrix $\underset{\sim}{R}_1 s_Y^2$. The variance of $\hat{y}_{(1)i} = \exp(\hat{Y}_{(1)i})$ is then given approximately by

$$V(\hat{y}_{(1)i}) = \text{var}(\hat{Y}_{(1)i})\{\hat{y}_{(1)i}\}^2 .$$

The losses of efficiency may be judged from the column in the table showing the ratios $\dfrac{\text{var } \hat{y}_{i(2)}}{\text{var } \hat{y}_{i(1)}}$. It will be seen that all are greater than 1, and that there are many very large values with one value as high as 308. We are thus in this example faced with a very serious loss of information that would result from using an inappropriate transformation, namely the original data.

In considering the results of Table 1, two influences should be borne in mind: (a) the parsimony effect, and (b) the effect of inappropriate weighting.

Concerning parsimony, the first order model contains four parameters; the second order model contains ten. It is well known that for any linear model containing p separately estimable parameters with n observations, irrespective of the design, the average variance of the estimated responses is $\frac{p}{n}\sigma^2$. Thus, associated with the use of the more parsimonious model, we should expect an average reduction in the variances of the fitted responses by a factor of $\frac{10}{4} = 2.5$.

Concerning inappropriate weighting, on the assumptions A_1 for the first order model in $Y = \log y$, the variances for the y_i will be heterogeneous, implying that weighted rather than unweighted least squares would be appropriate. It is well known that moderate heterogeneity of variances do not greatly affect estimates and their estimated standard errors, but in the case considered, this heterogeneity is extreme. For example, consider the ratio

$$\frac{\mathrm{var}(y_i)}{\mathrm{var}(y_j)} \doteq \left\{ \frac{\hat{y}_{(1)i}}{\hat{y}_{(1)j}} \right\}^2 .$$

In the most extreme case this is equal to

$$\frac{\mathrm{var}(y_{19})}{\mathrm{var}(y_9)} \doteq \left(\frac{3607}{88}\right)^2 = 41^2 = 1681 .$$

Thus given the appropriateness of the log metric, the variances for the y's differ by huge amounts and ordinary unweighted least squares will be very inefficient.

			Quadratic model in y		Linear model in $Y = \log y$	
			Perceived	Actual	Attainable	
x_1	x_2	x_3	$V(\hat{y}_{(2)})$	$V(\hat{y}_{(2)})$	$V(\hat{y}_{(1)})$	$V(\hat{y}_{(2)})/V(\hat{y}_{(1)})$
-	-	-	37.6	13.0	3.3	4.0
-	-	0	25.3	6.4	1.1	5.9
-	-	+	37.6	6.1	0.7	9.0
-	0	-	25.3	5.8	0.7	8.6
-	0	0	19.2	6.4	0.2	33.7
-	0	+	25.3	4.8	0.1	34.5
-	+	-	37.6	6.3	0.3	24.0
-	+	0	25.3	5.2	0.1	58.2
-	+	+	37.6	15.4	0.05	308.0
0	-	-	25.3	37.7	12.6	3.0
0	-	0	19.2	11.9	3.6	3.3
0	-	+	25.3	11.2	2.6	4.3
0	0	-	19.2	9.6	2.2	4.3
0	0	0	19.2	7.6	0.4	19.0
0	0	+	19.2	6.7	0.5	14.6
0	+	-	25.3	6.9	1.0	6.9
0	+	0	19.2	6.4	0.3	22.1
0	+	+	25.3	4.8	0.2	22.8
+	-	-	37.6	139.7	91.3	1.5
+	-	0	25.3	56.0	30.3	1.9
+	-	+	37.6	40.5	19.0	2.1
+	0	-	25.3	43.8	18.8	2.3
+	0	0	19.2	14.0	5.4	2.6
+	0	+	25.3	12.7	3.9	3.3
+	+	-	37.6	20.7	7.3	2.8
+	+	0	25.3	8.8	2.4	3.6
+	+	+	37.6	7.6	1.5	5.0

Table 1. Comparison of variances of estimated responses for 3^3 textile example.

(variances shown are $10^{-3} \times$ the actual variances)

5. DISCUSSION

Experience with the procedures proposed by Box and Cox confirms our belief that, when employed with data that potentially contain some useful information about transformation, these methods are not troublesome and can be extremely valuable. Theoretical support comes from the original work of Box and Cox and more recently from that of Hinkley and Runger (1983).

The alarming results of Bickel and Doksum concerning the supposed cost and instability of these methods follow largely from their failure to take account of arbitrary scaling (Box and Cox 1982). The high cost of not transforming data for which transformation is needed is illustrated by an example.

REFERENCES

BICKEL, P. J. and DOKSUM, K. A. (1981). An analysis of transformations revisited. J. Amer. Statist. Ass., 76, 296-311.

BOX, G. E. P. and COX, D. R. (1964). An analysis of transformations (with Discussion). J. Roy. Statist. Soc., B 26, 211-252.

BOX, G. E. P. and COX, D. R. (1982). An analysis of transformations revisited, rebutted. J. Amer. Statist. Ass., 77, 209-210.

CARROLL, R. J. and RUPPERT, D. (1980). On prediction and the power transformation family. Technical Report, University of North Carolina.

HINKLEY, D. V. and RUNGER, G. (1983). The analysis of transformed data. J. Amer. Statist. Ass., to appear.

GEPB:CF:scr

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER 2609	2. GOVT ACCESSION NO. AD-A137960	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle) SOME CONSIDERATIONS IN ESTIMATING DATA TRANSFORMATIONS		5. TYPE OF REPORT & PERIOD COVERED Summary Report - no specific reporting period
		6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s) George E. P. Box and Conrad A. Fung		8. CONTRACT OR GRANT NUMBER(s) DAAG29-80-C-0041
9. PERFORMING ORGANIZATION NAME AND ADDRESS Mathematics Research Center, University of Wisconsin, 610 Walnut Street, Madison, Wisconsin 53706		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS Work Unit Number 4 Statistics and Probability
11. CONTROLLING OFFICE NAME AND ADDRESS U. S. Army Research Office, P.O. Box 12211, Research Triangle Park, North Carolina 27709		12. REPORT DATE December 1983
		13. NUMBER OF PAGES 16
14. MONITORING AGENCY NAME & ADDRESS(if different from Controlling Office)		15. SECURITY CLASS. (of this report) UNCLASSIFIED
		15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

Approved for public release; distribution unlimited.

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

Transformations
Box-Cox
Bickel-Doksum
Efficiency
Stability

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

In a recent paper Bickel and Doksum claimed that procedures proposed by Box and Cox for estimating a transformation can be costly and unstable. We consider how the supposed cost and instability arise and illustrate our points by further analysis of textile data from the original paper. The analysis is used to make the further point that the cost of not making a transformation, when such is appropriate, can be extremely high. Some common sense advice on transformation analysis is given.

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

FILM

3-8